

**Exhibit 1**  
In Support of  
Non-Party Meta Platforms, Inc.'s  
Motion for Protective Order



KEN PAXTON  
ATTORNEY GENERAL OF TEXAS

**CIVIL INVESTIGATIVE DEMAND**

To: Facebook, Inc.  
c/o Jesse Solomon  
Davis Polk & Wardwell LLP  
901 15th Street, N.W.  
Washington, DC 20005  
[jesse.solomon@davispolk.com](mailto:jesse.solomon@davispolk.com)

The Office of the Attorney General of Texas (“OAG”) is investigating anticompetitive conduct in online markets, including markets relating to online advertising in Texas and the rest of the United States. Such activity may violate Section 15.05(b) of the Texas Free Enterprise and Antitrust Act of 1983, Tex. Bus. & Com. Code § 15.01 *et seq.* (the “Act”).

The OAG has reason to believe that Facebook, Inc (“Facebook”) may have information and/or material relevant to its investigation and is therefore issuing you this Civil Investigative Demand (“CID”) pursuant to Section 15.10 of the Act. Please review the following definitions and instructions carefully before responding to this CID. The documents you submit must be produced under a sworn certificate in the form attached as Exhibit A. Your response should be submitted **on or before November 6, 2020**, to Assistant Attorney General Margaret Sharp, Antitrust Division, Texas Office of the Attorney General, 300 W. 15th Street, 7th Floor, Austin, Texas 78701 or, if delivered by mail, at P.O. Box 12548, Austin, Texas 78711.

Section 15.10(i)(5) of the Act governs OAG's treatment of documents designated as containing trade secrets or confidential information. You should clearly designate documents or portions thereof, if any, that contain trade secrets or confidential information.

**NOTICE**

Any person who, with intent to avoid, evade, or prevent compliance, in whole or in part, with a CID, removes from any place, conceals, withholds, destroys, mutilates, alters, or by any other means falsifies any documentary material, or otherwise provides inaccurate information, is guilty of a misdemeanor which, on conviction, is punishable by a fine of not more than \$5,000, or by confinement in county jail for not more than one year, or both. Objections to this demand may be made in accordance with Section 15.10 of the Act.

Issued on October 5, 2020.

KEN PAXTON  
Attorney General of Texas

BRENT WEBSTER  
First Assistant Attorney General

RYAN L. BANGERT  
Deputy First Assistant Attorney General

DARREN McCARTY  
Deputy Attorney General for Civil Litigation

KIM VAN WINKLE  
Chief, Antitrust Division

*/s/ Margaret Sharp*

---

MARGARET SHARP  
Assistant Attorney General  
Antitrust Division  
(512) 463-1265  
(512) 320-0975 (fax)  
Margaret.Sharp@oag.texas.gov

**CERTIFICATE OF SERVICE**

I certify that on October 5, 2020, a true and correct copy of this Civil Investigative Demand was sent via certified mail, return receipt requested, and via electronic mail, addressed as follows:

Facebook, Inc.  
c/o Jesse Solomon  
Davis Polk & Wardwell LLP  
901 15th Street, N.W.  
Washington, DC 20005  
[jesse.solomon@davispolk.com](mailto:jesse.solomon@davispolk.com)

*/s/ Margaret Sharp*

---

MARGARET SHARP  
Assistant Attorney General

## DEFINITIONS

For the purpose of the request for documents, the following definitions apply.

1. “Document” means the original and all non-identical copies (whether different from the original because of notes, underlining, attachments, or otherwise), all computer files, and any written, printed, or recorded material of every kind in your possession, custody, or control. The term “computer files” includes information stored in or accessible through computers or other information retrieval systems. Thus, you should produce documents that exist in machine-readable form, including documents stored in personal computers, portable computers, handheld devices, workstations, minicomputers, mainframes, servers, backup disks and tapes, archive disks and tapes, and other forms of offline storage, whether on or off your premises. This definition covers electronic mail messages (“e-mail”), calendar entries, and other electronic documents in the possession of your employees, either at home or in their office.
2. “You,” “Your,” or the “Company” means Facebook, Inc., its successors, predecessors, divisions, wholly or partially owned subsidiaries, domestic or foreign parents, affiliates, partnerships, and joint ventures, and all directors, officers, trustees, employees, agents, and representatives of the foregoing.
3. “Demand Side Platform” or “DSP” means the technology and automated service that is used to purchase digital advertising inventory.
4. “Supply Side Platform” or “SSP” means mean the technology and automated service that is used to sell digital advertising inventory, including but not limited to real-time bidding auctions and other programmatic sales.
5. “Exchange” means the digital marketplace that enables the buying and selling of advertising space. The term includes, but is not limited to, advertising exchanges such as Facebook’s Ad Exchange and Google’s AdX.
6. “Ad Server” means the technology and service used to manage, or the process of managing, advertisers’ digital advertising campaigns, manage the sale of publishers’ digital advertising inventory, serve advertising creatives on digital properties, measure or track digital advertising campaigns, and provide reports on advertising campaigns (including attributions).
7. “AdTech Products” shall mean any and all systems and software, including predecessors and successors, used in the process of identifying, selecting, transmitting, and/or rendering display ads on a desktop or mobile device, as well as all systems and software, including predecessors and successors, that implement and/or effectuate an

Ad Auction, including, but not limited to, receiving, processing, and transmitting Bid Requests, Bid Responses, or related messages. The term includes but is not limited to Facebook Audience Network, Facebook Atlas, Facebook Ad Exchange, Facebook Ad Manager and Facebook LiveRail.

8. “FAN” means Facebook Audience Network.
9. “FBX” means Facebook Ad Exchange.

## INSTRUCTIONS

1. Except as otherwise specified, this CID requires production of Documents and information from January 1, 2014, to the present day.
2. This request is continuing in nature. To the extent that Documents or information responsive to this CID are obtained or generated after the issuance of this CID, Your production and responses should be supplemented accordingly.
3. For Your response to this CID to be complete, the certification form attached as **Exhibit A** must be executed by the official supervising compliance with this CID, notarized, and submitted along with the responsive materials and information.
4. For the purposes of the request for Documents, the following instructions apply:
  - A. No copying costs will be reimbursed unless a representative from the OAG has provided written approval, prior to submission, of the estimated cost of providing copies of the non-electronic documents requested in this CID. All other costs of complying with this CID will be borne by You.
  - B. If all or any portion of any responsive Document is withheld for any reason, including a claim of attorney-client privilege, any other applicable privilege or immunity, or judicial order, please submit a list of the documents withheld and state individually as to each document: (1) the source of the document; (2) the name, address, position and organization of the author and recipient of the document; (3) the type, title, specific matter, length and date of the document; and (4) the reason for withholding the document, including, but not limited to, identifying the actual or anticipated litigation that underlies a claim of work product immunity.
  - C. Documents produced electronically must comply with the Data Delivery Production Requirements attached hereto as **Exhibit B**.
  - D. All Documents responsive to this request for Documents, regardless of format or form and regardless of whether submitted in paper or electronic form:
    - a. shall be produced in complete form, unredacted unless privileged, and in the order in which they appear in Your files and shall not be shuffled or otherwise rearranged. For example:
      - i. if in their original condition papers were stapled, clipped, or otherwise fastened together or maintained in file folders, binders, covers, or containers, they shall be produced in such form, and any documents that must be removed from their original folders, binders, covers, or containers

to be produced shall be identified in a manner so as to clearly specify the folder, binder, cover, or container from which such documents came; and

- ii. if in their original condition electronic documents were maintained in folders or otherwise organized, they shall be produced in such form and information shall be produced so as to clearly specify the folder or organization format;
  - b. shall be produced in color where necessary to interpret the document;
  - c. shall be marked on each page with corporate identification and consecutive document control numbers;
  - d. shall be accompanied by an index that identifies: (i) the name of each person from whom responsive documents are submitted; and (ii) the corresponding consecutive document control number(s) used to identify that person's documents, and if submitted in paper form, the box number containing such documents. If the index exists as a computer file(s), provide the index both as a printed hard copy and in machine-readable form (provided that OAG representatives determine prior to submission that the machine-readable form would be in a format that allows the agency to use the computer files).
- E. If books and records that provide accurate responses are not available, enter best estimates and describe how the estimates were derived, including the sources or bases of such estimates. Estimated data should be followed by the notation "est." If there is no reasonable way for You to make an estimate, provide an explanation.
5. If documents responsive to a particular Document request no longer exist for reasons other than the ordinary course of business or the implementation of Your document retention policy as disclosed or described in response to a document request, but You have reason to believe have been in existence, state the circumstances under which they were lost or destroyed, describe the documents to the fullest extent possible, state the document request(s) to which they are responsive, and identify persons having knowledge of the content of such documents.
  6. If You are unable to respond to any request for Documents or information fully, explain why such response is incomplete, the efforts You made to obtain the information, and the source from which the complete response may be obtained.
  7. The terms "and" and "or" have both conjunctive and disjunctive meanings.

**Questions concerning this CID should be directed to Assistant Attorney General  
Margaret Sharp (512-463-1265, Margaret.Sharp@oag.texas.gov).**

## REQUESTS FOR DOCUMENTS

1. Produce a copy of all documents related to the consideration of Your Adtech products joining Google's Exchange Bidding and Open Bidding programs, including all internal correspondence, correspondence with Google, and studies underlying or otherwise relating to Your decision and Your evaluation of any alternatives.
2. Produce a copy of all documents related to the consideration of using Header Bidding solutions integrated with Your Adtech products and any alternatives considered, including all internal correspondence, correspondence with third parties related to these considerations, correspondence with Header Bidding providers, and studies underlying or otherwise relating to Your decision and Your evaluation of any alternatives.
3. Produce a copy of all documents that relate to Your decision to implement considered or actual launches or expansions of Your Adtech Products, including but not limited to, strategy documents, white papers, competitive impact studies, and market analyses preceding that decision as well as those evaluating the impact and performance of Your Adtech Products as a result of Your decision.
4. Produce a copy of all documents that relate to Your decision to discontinue, repurpose or contract the offerings from Your Adtech Products, including but not limited to, strategy documents, white papers, competitive impact studies, and market analyses preceding that decision as well as those evaluating the impact and performance of Your Adtech Products preceding and as a result of Your decision (if not discontinued).
5. Produce a copy of all documents related to adtech competitor analysis, competitive impact studies, and market analyses including but not limited to Google's products.
6. Describe and quantify how much revenue and impressions transacted through Your ad network product (Facebook Audience Network) correspond to ads shown on Your own and operated inventory versus third party inventory.
7. Produce a copy of all documents related to systems, efforts, experiments, or projects to assign identification numbers to users based on any method including but not limited to cookies or browser fingerprinting, to associate those identities with a profile storing information about users, to draw inferences about users by modelling their behavior, and/or to sync those identities and user profiles both across Your internal systems and with external partners for advertising purposes.
8. Produce a copy of all documents related to any analysis or discussion of decisions about acquisitions of companies or products, including but not limited to Atlas and LiveRail.
9. Produce a copy of all documents, answers to interrogatories, written questions, responses to Civil Investigative Demands, subpoenas, correspondence, and all other written material

which you receive[d] from or submit[ted] to the Department of Justice relating to its investigation into the business of Google. All documents produced should be identical to, in the same form as, and organized in the same manner as those produced to the Department of Justice.

10. State for each month from launch date to July 2020, for the following metrics regarding digital advertising separately for each of Your Own and Operated ad inventory in relation to the ad transactions and auctions that Facebook facilitates:
  - A. total number of impressions generated;
  - B. total number of impressions sold;
  - C. average number of bids per impression;
  - D. quartiles of number of bids per impression;
  - E. total revenues received from advertising sold through the ad server;
  - F. total publisher ad server fees paid.
11. For each metric in Request for Production #10 A-F, please provide the data broken down by:
  - type of transaction, including but not limited to header bidding, open/exchange bidding, programmatic guaranteed, private marketplace, open auction, first vs second price auction, and any other category maintained in the ordinary course of your business;
  - type of user device, including but not limited to desktop, mobile web, mobile application, and any other category maintained in the ordinary course of your business;
  - type of digital advertising, including but not limited to display, video, native, social, and any other category maintained in the ordinary course of your business; and
  - each SSP, ad exchange, or ad network transacting the impression; and
  - each advertiser ad server, or other source, bidding on the ad impression.
12. Produce the log-level data files for each impression, ad request, or bids between each of Your integrated Adtech products, each buyer/bidder sources and each ad inventory source from launch date to July 2020 for Your Own and Operated ad inventory.
13. State for each month from launch date to July 2020, for the following metrics regarding digital advertising Facebook Audience Network in relation to the ad transactions and auctions that Facebook:
  - A. total number of impressions on which FAN received a request to bid;
  - B. total number of impressions on which FAN entered a bid or requested a bid from advertisers;
  - C. total number of ad or bid requests that FAN bid into;
  - D. total number of bids received from advertisers;
  - E. total number of impressions transacted and won through FAN;

- F. the average and quartile measures of “Estimated action rates” and “Ad quality” from the impressions transacted through FAN;
  - G. the following revenue/fees for all impressions transacted through FAN:
    - gross revenues paid by the advertiser to FAN;
    - net revenue paid by FAN to their Own & Operated inventory and third parties;
    - total fees charged by FAN to advertisers, if applicable;
    - total fees charged by FAN to publishers, if applicable;
    - average cost per thousand impressions (or cost per click, or other relevant metric that is maintained in the ordinary course of business);
  - H. the total number of bid requests sent to demand sources (considering the integration to third party exchanges, if applicable);
  - I. the total number of bids received as a response to the bid requests sent;
  - J. the total value of the bids responding to the requests;
  - K. average number of bids received per impression, and quartiles of number of bids received per impression.
14. For each metric in Request for Production #13 A-K, please provide the data broken down by:
- type of transaction, including but not limited to header bidding, open/exchange bidding, programmatic guaranteed, private marketplace, open auction, first vs second price auction, and any other category maintained in the ordinary course of Your business;
  - type of user device, including but not limited to desktop, mobile web, mobile application, and any other category maintained in the ordinary course of your business;
  - type of digital advertising, including but not limited to display, video, native, social, and any other category maintained in the ordinary course of your business;
  - name of the buyer or bidder sources of the winning bid;
  - name of the source of the ad inventory, including Facebook O&O inventory and third-party in-app publishers and third-party web publishers;
  - whether the impression request included a cookie with an ID or not;
  - whether the impressions were matched on ID/cookies or not (corresponding to the case when a request is matched to FAN’s custom audience groups).
15. Produce the log-level data files Your Facebook Audience Network product for each impression, ad request, or bid request transacted between bidders/buyers and each ad inventory source from launch date to July 2020, including the values of “estimated action rates” and “ad quality”.
16. State for each month from January 2014 to latest available date, for the following metrics regarding digital advertising of your Ad Server product in relation to the ad transactions and auctions that Facebook facilitates:
- A. total number of impressions generated;

- B. the total number of impressions sold;
  - C. average number of bids received per impression;
  - D. quartiles of number of bids per impression;
  - E. total value of advertising sold each month;
  - F. net revenues received by publishers for advertising sold through the Ad Server;
  - G. the total fees charged to publishers for Ad Server usage.
17. For each metric in Request for Production #16 A-G, please provide the data broken down by:
- publisher-customers of the Ad Server product;
  - each party integrated with or connected to the Ad Server—whether by means of an application programming interface, software development kit, or otherwise—including but not limited to SSPs, DSPs, ad networks, and any other buyers or sellers of digital advertising;
  - type of transaction, including but not limited to header bidding, open/exchange bidding, programmatic guaranteed, private marketplace, open auction, and any other category maintained in the ordinary course of your business;
  - type of user device, including but not limited to desktop, mobile web, mobile application, and any other category maintained in the ordinary course of your business; and
  - type of digital advertising, including but not limited to display, video, native, social, and any other category maintained in the ordinary course of your business.
18. Produce the log-level data files for each impression, ad request, or bid request between the Ad Server and each exchange or SSP, or buyer/bidder, from January 2014 to July 2020 for Your Ad Server product.
19. State for each month from January 2014 to latest available date, for the following metrics regarding digital advertising of your Advertiser Ad Server product:
- A. total number of impressions purchased;
  - B. total value of advertising sold purchased month;
  - C. net revenues received by advertisers for advertising purchased through the Advertiser Ad Server;
  - D. the total fees charged to advertisers for Advertiser Ad Server usage.
20. For each metric in Request for Production #19 A-D, please provide the data broken down by:
- Advertiser-customers of the Advertiser Ad Server product;
  - each party integrated with or connected to the Ad Server—whether by means of an application programming interface, software development kit, or otherwise—including but not limited to SSPs, DSPs, ad networks, and any other buyers or sellers of digital advertising;

- type of transaction, including but not limited to header bidding, open/exchange bidding, programmatic guaranteed, private marketplace, open auction, and any other category maintained in the ordinary course of your business;
  - type of user device, including but not limited to desktop, mobile web, mobile application, and any other category maintained in the ordinary course of your business; and
  - type of digital advertising, including but not limited to display, video, native, social, and any other category maintained in the ordinary course of your business.
21. Produce the log-level data files for each impression and the ad auction that Facebook facilitates, ad request, or bid request between Your business and each ad inventory source, from January 2014 to July 2020 for Your Advertiser Ad Server product.
22. State for each month from launch date to latest available date, for the following metrics regarding digital advertising of FBX in relation to the ad transactions and auctions that Facebook facilitates:
- A. total number of impressions on which FBX received a request to bid;
  - B. total number of impressions on which FBX entered a bid or requested a bid from advertisers;
  - C. total number of ad or bid requests that FBX bid into;
  - D. total number of bids received from advertisers;
  - E. total number of impressions transacted and won through FBX;
  - F. the average and quartile measures of “Estimated action rates” and “Ad quality” from the impressions transacted through FBX, if applicable;
  - G. the following revenue/fees for all impressions transacted through FBX:
    - gross revenues paid by the advertiser to FBX;
    - total fees charged by FBX to advertisers, if applicable;
  - H. the total number of bid requests sent to demand sources (considering the integration to third party DSPs, if applicable);
  - I. the total number of bids received as a response to the bid requests sent;
  - J. the total value of the bids responding to the requests;
  - K. average number of bids received per impression, and quartiles of number of bids received per impression.
23. For each metric in Request for Production #22 A-K, please provide the data broken down by:
- type of transaction, including but not limited to header bidding, open/exchange bidding, programmatic guaranteed, private marketplace, open auction, first vs second price auction, and any other category maintained in the ordinary course of Your business;
  - type of user device, including but not limited to desktop, mobile web, mobile application, and any other category maintained in the ordinary course of your business;

- type of digital advertising, including but not limited to display, video, native, social, and any other category maintained in the ordinary course of your business;
  - name of the buyer or bidder sources of the winning bid;
  - whether the impression request included a cookie with an ID or not;
  - whether the impressions were matched on ID/cookies or not (corresponding to the case when a request is matched to FBX's custom audience groups).
24. Produce the log-level data files separately for each of Your Exchange product for each impression, ad request, or bid request between each of Your Exchange and each buyer/bidder sources and each ad inventory source from launch date to latest available date for Your Exchange product.
25. Produce data for each month from January 2014 date to July 2020 for the following metrics regarding the digital advertising of Your DSP business in relation to the ad transactions and auctions that Facebook facilitates:
- A. total number of ad or bid requests received;
  - B. total number of impressions on which the DSP requested bids from advertisers;
  - C. total number of ad or bid requests that Your DSP bid into;
  - D. total number of bids received from advertisers;
  - E. average number of bids received per impression, and quartiles of number of bids received per impression;
  - F. total number of impressions won by Your DSP;
  - G. following revenue/fees for all impressions offered and/or won by Your DSP:
    - gross revenues paid by the advertisers to Your DSP;
    - total fees charged by Your DSP to advertisers;
    - average cost per thousand impressions (or cost per click, or other relevant metric that is maintained in the ordinary course of business);
    - for third-party inventory in FAN purchased through Facebook Ad Manager, total fees charged to the advertiser and to the publisher, separately.
  - H. total value of the bids responding to the requests.
26. For each metric in Request for Production #25 A-H, provide the data broken down by:
- type of transaction, including but not limited to header bidding, open/exchange bidding, programmatic guaranteed, private marketplace, open auction, first vs second price auction, and any other category maintained in the ordinary course of Your business;
  - type of user device, including but not limited to desktop, mobile web, mobile application, and any other category maintained in the ordinary course of Your business;
  - type of digital advertising, including but not limited to display, video, native, social, and any other category maintained in the ordinary course of Your business; and
  - each advertiser ad server, or other source, requesting the ad impression.

- whether the impressions request included an ID
  - whether the impressions were matched on ID/cookies (corresponding to the case when a request is matched to the DSP's match table)
27. Produce the log-level data files for each impression, ad request, or bid request between Your business and each exchange or SSP, or ad inventory source, from January 2014 to July 2020 for Your DSP product in relation to the ad transactions and auctions that Facebook facilitates.

**EXHIBIT A**

**CERTIFICATE**

**STATE OF** \_\_\_\_\_

**COUNTY OF** \_\_\_\_\_

\_\_\_\_\_, being duly sworn upon his/her oath states:

I am a representative of Facebook, Inc. and have knowledge of the facts and circumstances relating to the collection and production of documents in response to this civil investigative demand. I hereby verify that all requested documents within the possession, custody, or control of Facebook, Inc. has been produced.

\_\_\_\_\_  
Title: \_\_\_\_\_

SUBSCRIBED and SWORN TO BEFORE ME this \_\_\_\_\_ day of \_\_\_\_\_, 2020.

\_\_\_\_\_  
NOTARY PUBLIC IN AND FOR THE  
STATE OF \_\_\_\_\_

## **EXHIBIT B**

### **DATA DELIVERY PRODUCTION REQUIREMENTS**

Prior to any production, a meeting with the Antitrust Division is required to discuss the location of any potentially relevant Electronically Stored Information (ESI) – to include discussing E-mail, Network Shares, Local Storage, Databases, Mobile Devices, the Cloud, Chat Applications, and Social Networking.

This document describes the standard specifications and procedures for making an image-based production to the Antitrust Division of the Office of the Attorney General of Texas (“OAG”) in the form of Load Files.

- To ensure the efficient processing and review of any electronic production, legal, technical, and economic OAG staff need to resolve the details prior to production, and preferably before you or your vendor begin to gather and process responsive documents.
- Care should be taken to ensure that all responsive data and metadata are preserved in the collection process.
- These are not Unicode compliant specifications and do not cover production of translations.
  - Please contact OAG staff if either of these is anticipated.

#### **I. Categories of Documents**

Discussion regarding the details of an electronic production should focus on seven categories of documents:

- A. Email and other electronic messages (e.g., instant messaging and chat applications)
- B. Other electronic documents
- C. Hard copy documents
- D. Shared resources
- E. Databases
- F. Audio and video data
- G. Foreign language materials

General requirements for each category of document are outlined below. For information regarding document-specific metadata and bibliographic information, please refer to the enclosed Metadata Table of Requested Fields.

#### **II. General Instructions for Productions**

To ensure the efficient processing and review of any electronic production, contact the OAG before you or your vendor begin to gather and process responsive documents.

- A. A cover letter should be included with each production of data. The following information should be included in the letter:
  - 1. List of each piece of media (hard drive, thumb drive, DVD or CD) included in the production by the unique number assigned to it, and readily apparent on the physical media.
  - 2. List of custodians, identifying:
    - a. The Bates range (and any gaps therein) for each custodian
    - b. Total number of records for each custodian
    - c. Total number of images for each custodian
    - d. Total number of native files for each custodian
  - 3. List of fields in the order in which they are listed in the data file
  - 4. Time zone in which emails were standardized during conversion (email collections only)
  - 5. Label all media with the following:
    - a. Case number, if applicable; if in response to an ID, the number of the ID
    - b. Volume Name
    - c. Production date
    - d. Bates range
    - e. Disk number (1 of X), if applicable
- B. Documents created or stored electronically MUST be produced in their original electronic format, not printed to paper or PDF.
- C. Data should be produced using whichever media requires the least number of deliverables.
- D. Organize productions by custodian, unless otherwise instructed. All documents from an individual custodian should be confined to a single load file.
- E. All productions should be checked and produced free of computer viruses.
- F. All produced media should be encrypted.
- G. Password-Protected Files: to the extent any produced documents are password-protected, the Producing Party must either unlock the document prior to production or provide passwords in order to allow access by the Receiving Party. If the Producing Party is unable to process a document because of unknown passwords or encryption, the Parties shall meet and confer to determine the appropriate steps to take with respect to such document.
- H. Production load files may also be electronically submitted via the internet (e.g., file share sites, FTP) Load Files in excess of 5GB should be discussed with the OAG in advance of submission.

### **III. Categories of Documents**

#### **A. Email, Attachments, and Other Electronic Messages**

Email and other electronic messages (*e.g.*, instant messages (IMs)) should be produced as image files with related searchable text, metadata and bibliographic information. Depending on how the company's systems represent names in email messages or IMs, a table of names or contact lists from custodians may be required.

Email repositories, also known as email databases (*e.g.*, Outlook .PST, Lotus .NSF), can contain a variety of items, including messages, calendars, contacts, tasks, etc. For purposes of production, responsive items should include the "Email" metadata/database fields outlined in the Metadata Table, including but not limited to, all parent items (mail, calendar, contacts, tasks, notes, etc.) and child files (attachments of files to email or other items), with the parent/child relationship preserved. Similar items found and collected outside an email repository (*e.g.*, .MSG, .EML, .HTM, .MHT) should be produced in the same manner.

Each IM conversation should be produced as one document.

Outlook (.PST) and Lotus Notes (.NSF) email files may be produced in native file format. Production of these native files should be discussed in advance with the OAG.

Electronic phone records should be discussed with the OAG in advance of production to determine the optimal format of production.

#### **B. Other Electronic Documents**

Electronic documents include word-processing documents, spreadsheets, presentations, and all other electronic documents not specifically discussed elsewhere. These documents should be produced as image files with related extracted text, metadata and bibliographic information and links to the original native file. All passwords and encryption must be removed from electronic documents prior to production. The native file documents should be named per the BEGBATES#.

Refer to the Metadata Table of Requested fields for required metadata fields.

1. The following file formats should be produced in their native forms:
  - a. For Word documents, .DOC and .DOCX
  - b. For Excel Spreadsheets, .XLS and .XLSX.
  - c. For PowerPoint Presentations, .PPT and .PPTX
  - d. For Microsoft Access Databases, .MDB and .ACCDB
  - e. For WordPerfect documents, .WPD
  - f. For Adobe Acrobat Documents, .PDF

## 2. Spreadsheets

Spreadsheets should be produced in native format (*e.g.*, as .XLSX files), with extracted text for the entire document, metadata, and bibliographic information. Provide a single placeholder image of the spreadsheet. The placeholder image must contain at a minimum the BEGDOC#, FILENAME, and FILEPATH. The Bates range for a spreadsheet should be a single number (*e.g.*, ABC-JD-00000001 – ABC-JD-00000001). The linked native file name should match the BEGDOC# with the appropriate file extension.

## 3. Presentations

Presentations should be produced in full slide image format along with speaker notes (which should follow the full images of the slides) with related extracted text, metadata, and bibliographic information. Presentations should also be produced in native format (*e.g.*, as .PPT files). The linked native file name should match the BEGDOC#, FILENAME, and FILEPATH.

## 4. Hidden Text

All hidden text (*e.g.*, track changes, hidden columns, hidden slides, mark-ups, notes) shall be expanded and rendered in the extracted text file and associated image file.

## 5. Embedded Files

All embedded objects (*e.g.*, graphical files, Word documents, Excel spreadsheets, .wav files) that are found within an electronic document shall be produced so as to maintain the integrity of the source document as a single document. For purposes of production, the embedded files shall remain embedded as part of the original source document. Internal documents that are embedded or referenced via hyperlinks and hyperlinked shared resources must be produced as separate, attached documents following the instructions in this appendix.

## 6. Image-Only Files

All image-only files (non-searchable PDFs, multi-page TIFFs, Snipping Tool screenshots, etc., as well as all other images that contain text) shall be produced with associated OCR text, metadata, and identifying information. Each production shall state the number of image-only files and non-imaged files produced.

## 7. Archive File Types

Archive file types (*e.g.*, .zip, .rar) must be uncompressed for processing. Each file contained within an archive file should be produced as a child to the parent archive file. If the archive file is itself an attachment, that parent/child relationship must also be preserved.

## 8. Proprietary File Types and Non-PC or Non-Windows Based Systems

Proprietary file types, such as those generated by financial or graphic design software, or file types from non-PC or non-Windows based systems (*e.g.*, Apple, UNIX, LINUX systems), should be discussed with the OAG in advance of production to determine the optimal format of production.

### C. Hard-Copy Documents

Hard-copy documents (*e.g.*, documents ordinarily maintained in paper) are to be produced as black-and-white image files, except where noted below, with related searchable OCR text and identifying information. Special attention should be paid to ensure that hard-copy documents are produced as they are kept, reflecting attachment relationships between documents and information about the file folders within which each document is found. In addition, multi-page documents must be produced as single documents (*i.e.*, properly unitized) and not as several single-page documents. Where color is required to interpret the document, such as hard copy photos, and certain charts, its image MUST be produced in color. These color images are to be produced as JPEG format. Hard-copy photographs should be produced as color JPEGs, if originally in color, or grayscale TIFF files if originally in black-and-white.

### D. Shared Resources

Shared Resources should be produced as separate custodians if responsive custodians have access to them or if they contain responsive documents. The name of the group having access would be used as the custodian name, *i.e.*, Marketing Execs or Accounting Dept. The company will separately provide a brief description of each shared resource that includes a list of the custodians who have access to that shared resource.

### E. Database Productions

Production of enterprise databases are not addressed in these specifications and must be discussed with the appropriate legal and technical staff to determine the optimal production format; these will usually fall outside the scope of an image-based production. Care must be taken to ensure that all responsive databases and their metadata are preserved.

### F. Audio and Video Data

Audio and video file types not specifically discussed elsewhere should be discussed with the OAG in advance of production to determine the optimal format of production.

### G. Foreign Language Materials

Production of Foreign Language Materials not addressed in these specifications must be discussed with the appropriate legal and technical staff to determine the optimal production format prior to production.

## IV. Processing

### A. De-Duplication

Before doing any de-duplication, provide the OAG with a written description of the method that will be used to de-duplicate (including which elements are compared and what hash codes are used), and what is considered a duplicate. Then confirm that your approach is acceptable to the OAG. De-duplication of hard-copy documents, or that of “loose” electronic documents (*e.g.*, presentation slides located on the custodian’s C: drive) against email attachment versions of those same documents, is unacceptable. The integrity of any produced email and any related “document family” must be maintained except as limited by any claim of privilege. De-

duplication should occur both vertically within each custodian and horizontally across custodians. Vertical de-duplication is crucial when a production includes electronic documents from back-ups. Horizontal de-duplication must be done in a way that preserves (and produces) information on blind copy (Bcc) recipients of emails and other custodians whose files contain the duplicates that will be eliminated from the production.

A Custodian Append File is to be produced when de-duplicating ACROSS custodians (*i.e.*, horizontal de-duplication) and data is produced on a rolling basis. The file must be provided on an incremental basis **starting with the second submission**; as more custodians are discovered for previously produced documents, this file is updated with **only the new** custodian information.

The Custodian Append File is a two-field delimited file consisting of the BEGDOC#s of the previously delivered document and the **new** custodian names for the duplicates of those records that would otherwise have been produced in the subsequent (new) submissions.

These specifications do not allow for near de-duplication or email threading. These formats must be discussed separately with the OAG and consent obtained prior to the use of such techniques for production.

#### B. Document Numbering

Documents must be uniquely and sequentially Bates-numbered across the entire production, with an endorsement burned into each image. Each Bates number shall be of a consistent length, include leading zeros in the number, and unique for each produced page. Bates numbers should contain **no more** than three segments. For example, a company identifier, a middle segment identifying the custodian, and a sequential page counter with connecting hyphens. The number of digits in the numeric portion of the format should not change in subsequent productions, nor should spaces, hyphens, or other separators be added or deleted. Under no circumstances should bates numbers contain embedded spaces, slashes (/), backslashes (\), carats (^), ampersands (&), hash marks (#), plus signs (+), percent signs (%), dollar signs (\$), exclamation marks (!), pipes (|), any character used as a delimiter in the metadata load files, or any character not allowed in Windows file-naming convention (,\ / : \* ? “ < > | ~ @ ^). Bates numbers may contain hyphens (-).

#### C. Privilege Designations

Documents redacted pursuant to any claim of privilege will be designated “Redacted” in the SPECPROPERTIES field as described in the Metadata Table. Appropriately redacted searchable text (OCR of the redacted images is acceptable), metadata, and bibliographic information must also be provided. All documents that are part of a document family that includes a document withheld pursuant to any claim of privilege will be designated “Family Member of Privileged Doc” in the SPECPROPERTIES field as described in the Metadata Fields table for all other documents in its family. Placeholder images with BEGDOC#, FILENAME, FILEPATH and reason withheld (*e.g.*, “Privileged”) should be provided in place of the document images of the privileged document.

## IMAGE AND TEXT FILE SPECIFICATIONS, AND LOAD FILE CONFIGURATION

### **I. Load File Configuration**

- Each production must have a unique VOLUMENAME (e.g., TXAG\_VOL001)
- The VOLUMENAME should increase sequentially with each subsequent production (e.g., TXAG\_VOL001, TXAG\_VOL002, etc.)
- Under the VOLUMENAME folder, the production should be organized into 4 subfolders:
  - DATA (should contain metadata, DAT, OPT, custodian append, etc. files)
  - IMAGES (may contain subfolders, with no more than 5,000 images per folder)
  - NATIVES (may contain subfolders, containing linked native files, with no more than 5,000 files per folder)
  - TEXT (may contain subfolders, with document-level text files)

### **II. Image/Native File Specifications**

- Black-and-white Group IV Single-Page TIFFs (300 DPI). Color images should be provided in .jpg format when color is necessary.
- Image file names should match the page identifier for that specific image and end with the .tif (or .jpg if needed) extension.
  - Example: ACME-ABC-0003072.TIF
- File names cannot have embedded spaces, commas, ampersands, slashes, back slashes, hash marks, plus signs, percent signs, exclamation marks, any character used as a delimiter in the metadata load files, or any character not allowed in Windows file-naming convention. (, & \ / # + % ! : \* ? “ < > | ~ @ ^ )
- Images for a given document must reside together in the same folder.
- The maximum number of image files should be limited to 5,000 per folder.
- Native file names should match the BEGDOC entry for that specific record and end with the appropriate file extension.
- The maximum number of native files in a subfolder should be limited to 5,000 per folder.
- Any encryption or password protection will be removed from all native format files produced.

### **III. Extracted or Searchable Text File Specifications and Control List Configuration**

- Extracted text, as opposed to OCR, must be provided with all records, except for documents that originated as hard copy or redacted documents.
  - For hard copy documents, provide OCR text.
  - For redacted documents, provide OCR text for the redacted version.
- There should be a single extracted/OCR text file per document, in ASCII Text format only.

- The name of the text file should be the same as the document's first page/Bates number, with a TXT extension: BEGDOC#.TXT.
- There must be a carriage return and line feed (CRLF) no later than the 250th character of the first line of every text file.
- All soft and hard returns in the native electronic or image file should be replicated as a Carriage Return Line Feed (CRLF) in the text file (i.e., the lines of text in the file terminate with a CRLF in correlation with the appearance of the native electronic or rendered image file). Pay particular attention to not allow multi-line paragraphs of e-mails to be rendered as a single, extended line of text.
- Text files should include page breaks that correspond to the “pagination” of the image files. Place text files under a “TEXT” folder and provide a Control List file for loading in the “DATA” folder on the delivery media.

#### **IV. Image Load File (.opt) Configuration**

- Page level comma-delimited file containing seven fields per line.
  - PageID,VolumeLabel,ImagePath,DocumentBreak,FolderBreak,BoxBreak,PageCount
    - PageID – PageID of the item being loaded. MUST be identical to the image name (less the file extension).
    - VolumeLabel – Optional. If used it is preferable that it match the VOLUMENAME assigned in the corresponding metadata load file.
    - ImageFilePath – The path to the image from the root of the delivery media.
    - DocumentBreak – The letter “Y” denotes the first page of a document. If this field is blank the page is not the first page of a document.
    - FolderBreak – Leave empty
    - BoxBreak – Leave empty
    - PageCount – Optional
- Example - ABC-JD00030005,,\ABC002\Images\001\ ABC-JD-00030005.tif,Y,,,

#### **V. Metadata Load File Delimiters and Configuration**

- Field Separator ¶ (ASCII 020)
- Text Qualifier ¢ (ASCII 254)
- Substitute Carriage Return or New Line in data ® (ASCII 174)
- Multi-value separator (Do Not Follow with Space) ; (ASCII 059)
- Date format YYYY.MM.DD (date type fields only)
- Time format HH:MM:SS in 24-hour format (e.g., 04:32 pm formatted to 16:32:00 – Do not include AM, PM, or Timezone indicators).
- There should be one line for every record in the load file. A carriage return and line feed (CRLF) must appear at the end of each record and ONLY at the end of each record.
- The first row of each metadata load file should be a header row containing the field names. Field names must match OAG Metadata Table field names.
- All requested fields should be present in the metadata load file whether data exists or not. Field order must remain consistent in subsequent productions.

## VI. METADATA TABLE OF REQUESTED FIELDS

Field Name	Field Description	Field Type
VOLUMENAME	Production volume number (e.g., ABC001-001)	Note Text
BEGDOC	Start Bates (including prefix) -- No spaces or special characters	Note Text
ENDDOC	End Bates (including prefix) -- No spaces or special characters	Note Text
ATTACHRANGE	Range of the BEGDOC value of the parent record to the ENDDOC value (including prefix) of the last child record (for example, ABC-JD-00001201 ABC-JD-00001220); populated for all documents in the group. Empty if the record is NOT in family grouping	Note Text
BEGATTACH	Starting Bates of attachment range	Note Text
ENDATTACH	Ending Bates of attachment range	Note Text
PARENTBATES	Parent record's BEGDOC#, including prefix (populated ONLY in child records)	Note Text
CHILDBATES	Child document list: BEGDOC# of each child, separated by semicolons [:] (populated ONLY in parent records)	Multi-Entry
NUMPAGES	Page count	Integer
SPECPROPERTIES	Indicate all that apply : Record Type: E-Doc, E-Doc Attachment, Email, Email Attachment, Hard Copy, Calendar Appt Other Notations: Translation of [DOCID of original], Translated as [DOCID of Translation] Privilege Notations: Redacted, Privileged, Family Member of Priv Doc	Multi-Entry
CUSTODIAN	Custodian(s) / source(s) -- format: Last, First or ABC Dept	Multi-Entry
TIMEZONE	The time zone from which the native file was collected.	Note Text
FROM	Author of the Email or Calendar item (as formatted on the original)	Note Text
TO	Recipients of the Email (as formatted on the original)	Multi-Entry
CC	Names of the individuals who were copied on the Email (as formatted on the original)	Multi-Entry
BCC	Names of the individuals who were blind-copied on the Email (as formatted on the original)	Multi-Entry

Field Name	Field Description	Field Type
SUBJECT	Email, document, or calendar subject	Note Text
DATESENT	Date the Email was sent -- Format: YYYYMMDD.	Date
TIMESENT	Time Email was sent -- Format: HH:MM:SS	Time
DATERECEIVED	Date Email was received -- Format: YYYYMMDD	Date
TIMERECEIVED	Time Email was received -- Format: HH:MM:SS	Time
HEADER	The internet header information for Email sent through the internet	Note Text
INTERNETMSGID	Globally unique identifier for a message which typically includes messageid and a domain name. Example: <0E6648D558F338179524D555@m1p.innovy.net>	Note Text
MESSAGEID	Proprietary email database/mailstore/post office file associated with centrally managed enterprise email servers. Microsoft Outlook PST EntryID, the UniqueID (UNID) for Lotus Notes, equivalent value for other proprietary mailstore formats.	Note Text
CONVERSATIONINDEX	Email Thread Identification	Note Text
IMPORTANCE	Email flag indicating priority level set for message	Note Text
DATECREATED	Date electronic file was created -- Format: YYYYMMDD	Date
REVISION	Revision number extracted from metadata of electronic documents	Note Text
DATESAVED	Date native file was last modified; Format: YYYYMMDD.	Date
ORGANIZATION	Company field extracted from the metadata of an electronic document	Note Text
KEYWORDS	List of keywords, tags, and categories extracted from the metadata of an electronic document	Multi-Entry
AUTHOR	Author field value extracted from the metadata of a native file	Note Text
MODIFIED	Date of last modification as extracted from metadata of electronic documents -- Format: YYYYMMDD	Date
SPEC#	Subpoena/request paragraph number to which the document is responsive	Multi-Entry
SEARCHVALUES	List of search terms used to identify record as responsive (if used)	Multi-Entry
HASHMD5	Document MD5 hash value (used for de-duplication or other processing) NOTE: only one hash value metadata field need be provided—either MD5 hash or SHA1 hash.	Note Text

Field Name	Field Description	Field Type
HASHSHA	Document SHA1 hash value (used for de-duplication or other processing) NOTE: only one hash value metadata field need be provided—either MD5 hash or SHA1 hash	Note Text
FILESIZE	File size in Bytes (integer value only - do not include unit of measure or decimal places - e.g., 568)	Integer
FILENAME	File name of electronic document with the <b>file extension</b> (e.g., Re: tomorrow sounds fun.eml; Read this today.docx)	Note Text
APPLICATION	Application used to create electronic document (e.g., Excel, Outlook, Word) as reflected in metadata	Note Text
FILEEXTENSION	File extension of electronic document (e.g., .eml, .msg, .xlsx)	Fixed Length 5 chars
FOLDERLABEL	Email folder path (sample: Inbox\Active); or Hard Copy folder/binder title/label	Note Text
FILEPATH	File path to electronic document as it existed in its original environment	Note Text
NATIVELINK	File path location to the current native file location on the delivery medium	Note Text